Ensemble of Neural Network Emulations for Climate Model Physics: The Impact on Climate Simulations

Michael S. Fox-Rabinovitz, Vladimir Krasnopolsky, and Alexei Belochitski

Abstract—**A new application of the NN ensemble approach is presented. It is applied to NN emulations of model physics in complex numerical climate models, and aimed at improving the accuracy of climate simulations. In particular, this approach is applied to NN emulations of the long wave radiation of the widely used National Center for Atmospheric Research Community Atmospheric Model. It is shown that practically all individual neural network emulations that we have trained in the process of development an optimal NN LWR emulation can be used within the NN ensemble approach for climate simulation. Using the NN ensemble results in a significant reduction of climate simulation errors, namely: the systematic and random errors, the magnitudes of the extreme errors or outliers and, in general, the number of large errors.**

I. INTRODUCTION

A variety of important practical applications of neural networks (NN) in geosciences [2-5, 10], including those of the numerical climate model components considered in this paper, may be treated mathematically as a mapping between two vectors X (input vector) and Y (output vector) and symbolically can be written as:

$$Y = M(X); \quad X \in \Re^n, Y \in \Re^m \qquad (1)$$

The simplest multi-layer perceptron (MLP) neural network (NN) is a generic analytical nonlinear approximation or model for mapping [1], like the mapping (1).

In the context of our application, we developed NN emulations of climate model physics [5] for the widely used National Center for Atmospheric Research (NCAR) Community Atmospheric Model (CAM). Specifically, a number of the NN emulations of the original long wave radiation (LWR) (mapping like (1)), have been individually trained, with slightly different approximation and interpolation accuracies. In this study, we investigate the ability of NN ensembles, created from these NN emulations, to provide a better approximation and interpolation than their individual members and, most importantly, a better accuracy of climate simulation.

As a nonlinear model or nonlinear approximation, the NN approximation problem allows for multiple solutions or for

multiple NN emulations of the same LWR. For example, the original LWR, used in NCAR CAM, can be approximated with NNs with different numbers of hidden neurons, with different weights (resulting from the NN training with different initializations), different partitions of the training set, etc. At the same time, these multiple NNs may be different in terms of other criteria providing complementary information about the target mapping. The availability of multiple NN emulations, providing additional/complimentary information about the target mapping, opens an attractive opportunity of introducing an ensemble approach. It allows for integrating the complimentary information, contained in the individual ensemble members, into an ensemble that "knows" more about or represents the original LWR better than each of the individual ensemble members (a particular NN emulation). Moreover, the NN ensemble, when it is used in a climate model instead of a single NN emulation of the original LWR, is expected to provide a better accuracy of the climate simulation.

An ensemble of NNs consists of a set of members, which are individually trained NNs. They are combined when applied to a new input data to improve the generalization (interpolation) ability. The previous research has shown that an ensemble is often more accurate than any or most of the individual ensemble members. Different ways of combining NN ensemble members into the ensemble have been developed and investigated [7]. In this work, we used a conservative ensemble [15] where simple (with equal weights for all members) averaging of the members provides the ensemble mean and other statistics.

The previous research also suggests that any mechanism that causes some randomness in or perturbation for the formation of NN ensemble members, can be used to form an accurate NN ensemble [8]. For example, ensemble members can be created by training different members: (a) on different subsets of the training set [8]; (b) on different sub-domains of the training domain; (c) using NNs with different topology (different number of hidden neurons) [9]; (d) using NNs with the same architecture but with different initial conditions for NN weights [10,11].

Most of the previous studies with NN ensembles have been done in the context of solving classification [10,13] or prediction of time series problems [7,11]. Also, the NN ensemble technique has been recently applied to improve the accuracy of the NN Jacobian and NN emulations for an

oceanic data assimilation application [12].

In section 2 of this paper we discuss an application of the NN ensemble approach to emulating NCAR CAM LWR. Conclusions are presented in section 3.

II. APPLICATION OF THE NN ENSEMBLE APPROACH TO THE NCAR CAM LONG WAVE RADIATION

Here we will describe an application of the NN ensemble approach in the context and as an extension of our development of NN emulations for the climate model physical processes [4-6] we worked on for the last several years. In this study, we use as a test bed the NCAR CAM LWR (see [14] and the references therein). NCAR CAM is a state-of-the-art widely recognized and used climate model. Our approach to developing NN emulations for NCAR CAM LWR is described in detail in [5].

A. NCAR CAM Long Wave Radiation

The function of the LWR in atmospheric GCMs is to calculate heating fluxes and rates produced by LWR processes. This LWR is physically and computationally very complex that makes it a major computational "bottleneck" in the NCAR CAM physics. The method for calculating LWR in the NCAR CAM is based on the long-wave radiative transfer equations in an absorptivity/emissivity formulation (see [14] and the references therein).

The input vectors for the NCAR CAM LWR include ten vertical profiles (atmospheric temperature, humidity, ozone, CO_2, N_2O, CH_4, two CFC mixing ratios (the annual mean atmospheric mole fractions for halocarbons), pressure, and cloudiness) and one relevant surface characteristic (the upward LWR flux at the surface). The CAM LWR output vectors consist of the vertical profile of heating rates (HRs) and several radiation fluxes, including the outgoing LWR flux from the top layer of the model atmosphere (the outgoing LWR or OLR).

The NN emulation of the NCAR CAM LWR has the same number of inputs (a total of 220) and outputs (a total of 33) as the original NCAR CAM LWR. In the process of development of an optimal NN emulation for the LWR, we have trained the multiple emulating NNs that all have one hidden layer with 20 to 500 hidden neurons (the ensemble-2 below). For some topologies we trained a set of emulating NNs with different initial conditions for obtaining NN weights. For example, we trained fourteen different NN emulations with 150 hidden neurons (the ensemble-1 below). Varying the number of hidden neurons and initial conditions when calculating the NN weights, allowed us to demonstrate the dependence of the accuracy of approximation as well as its convergence on these parameters [5,6]. As a result, an optimal NN emulation with the accuracy of approximation sufficient for a decadal integration in the climate model has been obtained. However, the remaining/intermediate NN emulations also contain potentially valuable information about the CAM LWR. In this paper, by using an NN ensemble, we used the majority of these intermediate NN emulations to improve the

quality of approximation and, most importantly, of generalization (interpolation) provided by our LWR NN emulations.

NCAR CAM is integrated for two years to generate representative data sets. The first year of the model simulation is divided into two independent parts, each containing input/output vector combinations. The first part is used for training and the second one for validation (control of overfitting, control of a NN architecture, etc.). The second year of model simulation is used for creating a test data set, completely independent from both training and validation data sets. This independent data set is used for testing only. All approximation statistics presented in this section are calculated using this independent test data set. These statistics illustrate the accuracy of the interpolation provided by NN emulations. All NN emulations and ensembles of the NN emulations are tested against the control which within our framework is obviously the original NCAR CAM LWR. Mean difference B (bias or a systematic error) and the root mean square difference $RMSE$ (a root mean square error) between the original LWR and its NN emulation, maximum and minimum errors, and distributions of errors over the entire test, are calculated. We used a conservative ensemble [15] with a simple (with equal weights for all members) averaging of the members providing the ensemble statistics.

The final and the most important test is performed by estimating the accuracy of decadal climate simulation runs with single NN ensemble members and with the NN ensemble vs. the control climate run with the original NCAR CAM LWR.

B. Ensemble-1

The first NN ensemble presented in this paper (ensemble-1) is a set of 12 emulating NNs with 150 neurons in one hidden layer that have been trained with different initial conditions for NN weights. Fig. 1 illustrates the spread (diversity) of the NN ensemble members in terms of the systematic and random interpolation errors. It shows the effectiveness of the ensemble approach in reduction of the systematic interpolation error (bias, the horizontal axis) and the random interpolation error (error standard deviation, the vertical axis), calculated on the independent test set. The ensemble bias is equal to the average of member's biases (because we use a conservative ensemble); however, the ensemble random error is smaller than the smallest member's error. Fig. 2 demonstrates a significant effectiveness of the ensemble mean in reducing the magnitude of extreme errors or outliers. The ensemble mean reduces maximum (vertical axis) and minimum (horizontal axis) errors by 2-3 times as compared to those of the worst ensemble member.

Fig.3 shows that not only the magnitude of the largest errors or extreme outliers is effectively reduced by the ensemble but the entire error distribution is changed. Fig. 3 shows the tail of the distribution of errors. It demonstrates that the number of larger errors (> 5 K/Day) for the
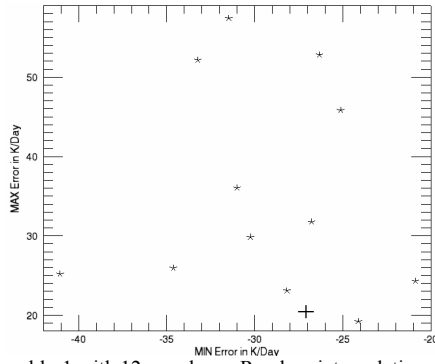
Fig.1. Ensemble-1 with 12 members. Random interpolation error (error SD) vs. systematic error (Bias). Asterisks – ensemble members, cross – ensemble.
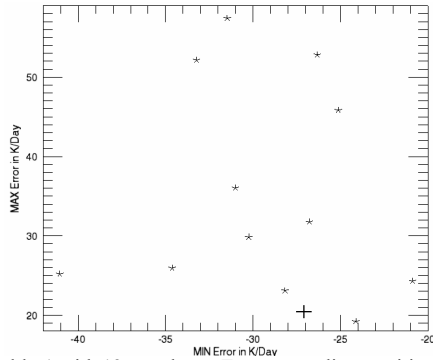

Fig. 2 Ensemble-1 with 12 members. Extreme outliers positions, maximum error vs. minimum error. Asterisks – ensemble members, cross – ensemble.
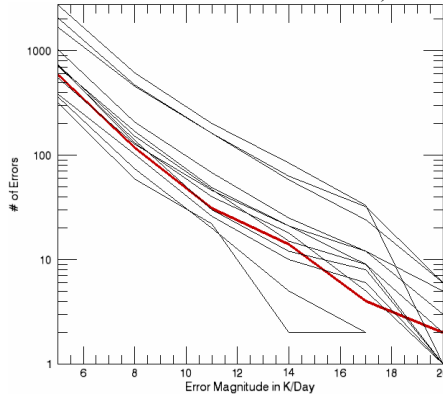

Fig. 3 Ensemble-1 with 12 members. Tails of the error distributions for ensemble members (thin solid, black lines) and for the ensemble (thick solid, red line). The logarithm of the number of errors (vertical axis) with respect of the magnitude of the error (horizontal axis). # of Errors correspond to the number of points with the errors of a magnitude indicated on the x-axis.

ensemble is almost an order of magnitude smaller than that for the worst ensemble member.

C. Ensemble-2

The second ensemble (ensemble-2) is composed of 14 members including emulating NNs that all have one hidden layer with the different number of hidden neurons, from 20 to 500. Figs. 4-6 contain info similar to that of Figs. 1-3 but presented for the ensemble-2. The comparison of these two sets of figures shows that the ensemble-2 is quite effective in reducing the errors for the ensemble mean but a bit less effective in this respect than the ensemble-1.
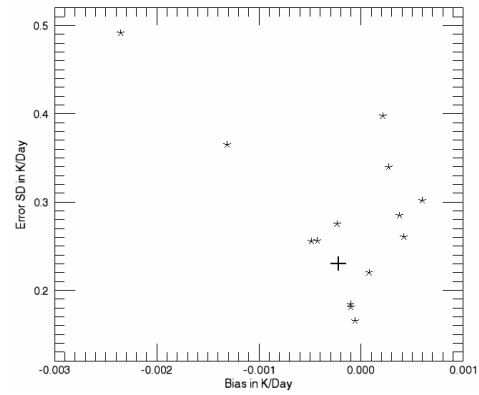

Fig. 4 Ensemble-2 with 14 members. Random interpolation error (error SD) vs. systematic error (Bias). Asterisks – ensemble members, cross – ensemble.
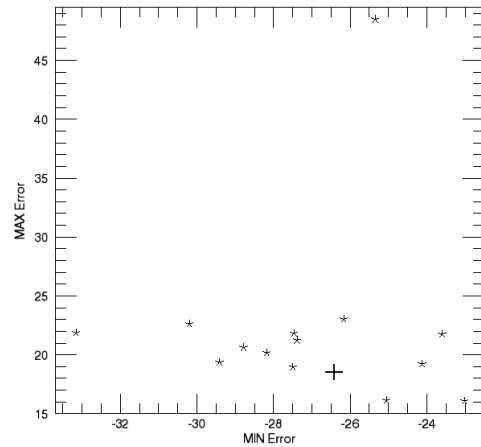

Fig. 5 Ensemble-2 with 14 members. Extreme outliers positions, maximum error vs. minimum error. Asterisks – ensemble members, cross – ensemble.

D. Ensemble-3

So far, we presented the errors of NN emulations calculated against the NCAR CAM original LWR on the test dataset. The next and most important step is validation of the NN ensemble approach for NCAR CAM long-term decadal climate simulations. This step allows us to arrive at an overall conclusion on the practical efficiency of the NN ensemble approach for climate simulations.

We have selected a sufficiently diverse group of six NN emulations from the ensembles-1 and-2. These six members constitute the ensemble-3. For the ensemble-3, the statistics calculated for the ensembles-1 and-2 on the test data set are also calculated. In addition, climate simulations have been run with NCAR CAM for 25 years with each of these six ensemble members (each of the six LWR NN emulations).

As it is usually done for climate simulations, the first 10 year simulated fields, that potentially include the climate model spin-up effects, are not used for the analysis of the simulation results so that the remaining 15 year period is used for the purpose. The results (climate fields and diagnostics) of each simulation are compared with the control climate run of NCAR CAM performed with the original LWR. The climate simulation errors (systematic,
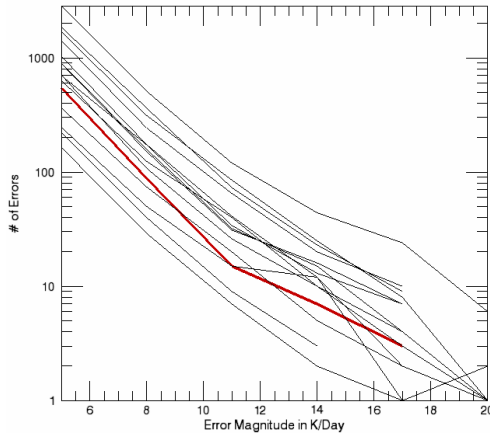
Fig. 6 Ensemble-2 with 14 members. Tails of the error distributions for ensemble members (thin solid, black lines) and for the ensemble (thick solid, red line). The logarithm of the number of the errors (vertical axis) with respect of the magnitude of the error (horizontal axis).
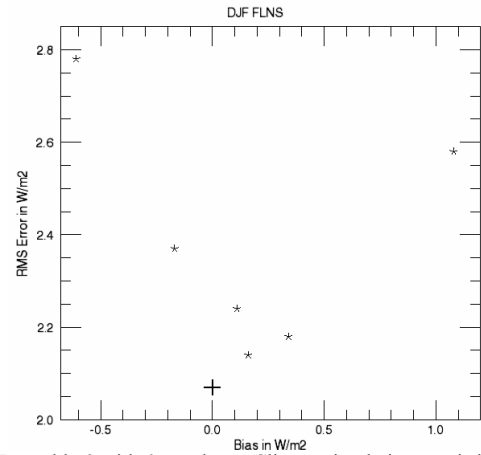

Fig. 7 Ensemble-3 with 6 members. Climate simulations statistics: RMS interpolation error vs. systematic error (Bias) for FLNS (see text). Asterisks – ensemble members, cross – climate NN ensemble results; DJF stands for December-January-February (winter).
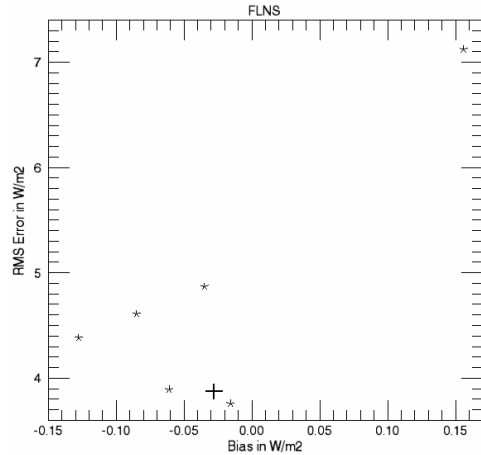

Fig. 8 Ensemble-3 with 6 members. RMS interpolation error vs. systematic error (Bias) for FLNS (see text) on independent test set. Asterisks – ensemble members, cross – NN ensemble.

random, maximum, and minimum) have been calculated for each ensemble member. Then the NN ensemble climate run has been performed. For this run, six NN emulations are applied and the LWR outputs are calculated as the mean of these six NN emulation outputs, at each time step and at each grid point throughout the model integration. The results of this NN ensemble climate run are shown by crosses in the following figures.

For consistency with the discussions in sub-sections 2.A-C, we present the validation statistics using the independent test set for the ensemble-3. We also present the validation statistics for ensemble-3 for the 15-year climate simulation. Figs. 7 – 10 show the statistics for the net surface LWR flux (FLNS in W/m²). Figs. 7 and 9 show the time averaged (15-year mean) errors for the NCAR CAM runs, with LWR NN emulations, calculated against the control run with the original LWR. Figs. 8 and 10 show the errors calculated on the independent test set used in sections *A* and *B*. Figs. 7 and 8 show the spread (diversity) of the NN ensemble members in terms of the systematic and RMS interpolation errors (in the case of small bias the RMS error is almost equal to the random error). We see the effectiveness of the NN ensemble approach in reduction of both the systematic interpolation error (bias, the horizontal axis) and the RMS interpolation error (the vertical axis) for the climate simulation and test set calculations. The NN climate ensemble (Fig. 7) bias is very small; the ensemble's random errors are smaller than the smallest member's error. For the NN ensemble errors for the test set (Fig. 8), the results are close to that of Fig. 7.

Figs. 9 and 10 show similar comparisons for the magnitude of the largest errors or extreme outliers. The NN ensemble effectively reduces these extreme errors. Figs. 11 – 14 show the same comparison statistics for the net LWR flux at the top of the model atmosphere (FLNT in W/m²). The FLNT results show the effective reduction of errors similar to that of FLNS. It is noteworthy that FLNT or OLR is well correlated with precipitation so that the positive impact from using the NN ensemble on FLNT is reflected in precipitation.

The comparisons of statistics estimated on an independent test set and for climate simulations show that the relative improvements in the accuracy of climate simulations are similar to (or even better than) those estimated on an independent test set. The use of the NN ensemble in climate simulation significantly reduces the systematic error (bias); it also reduces the random error to the value smaller than that of the best individual ensemble member. The same true for the extreme errors.

III. CONCLUSION

In this paper, we have presented a new application of the NN ensemble approach. We have applied the NN ensemble approach to improve the accuracy of climate simulations that use NN emulations of the model physics [5]. In particular, we applied this technique to NN emulations we developed for the LWR of NCAR CAM. We have shown that practically all individual NN emulations that we have trained in the process of development of an optimal NN emulation for LWR, can be used, within the NN
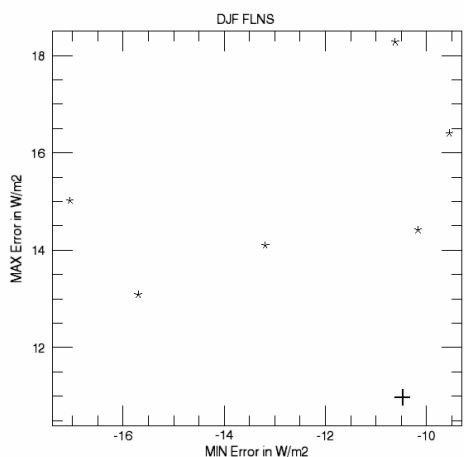
Fig. 9 Ensemble-3 with 6 members. Climate simulations statistics: extreme outliers positions, maximum error vs. minimum error for FLNS (see text). Asterisks – ensemble members, cross – climate NN ensemble results; DJF stands for December-January-February (winter).

Fig. 12 Ensemble-3 with 6 members. RMS interpolation error vs. systematic error (Bias) for FLNT (see text) on independent test set. Asterisks – ensemble members, cross – NN ensemble.
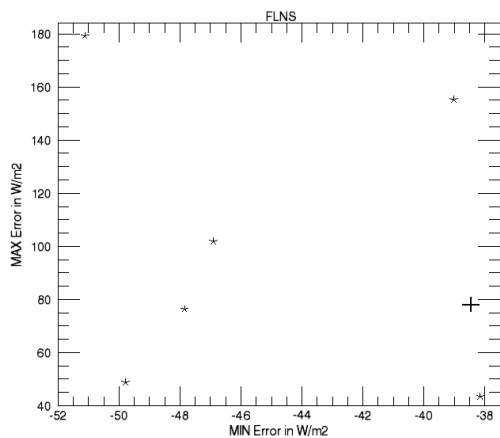

Fig. 10 Ensemble-3 with 6 members. Extreme outliers positions, maximum error vs. minimum error for FLNS (see text) on independent test set. Asterisks – ensemble members, cross – NN ensemble.

Fig. 13 Ensemble-3 with 6 members. Climate simulations statistics: extreme outliers positions, maximum error vs. minimum error for FLNT (see text). Asterisks – ensemble members, cross – climate NN ensemble results; DJF stands for December-January-February (winter).
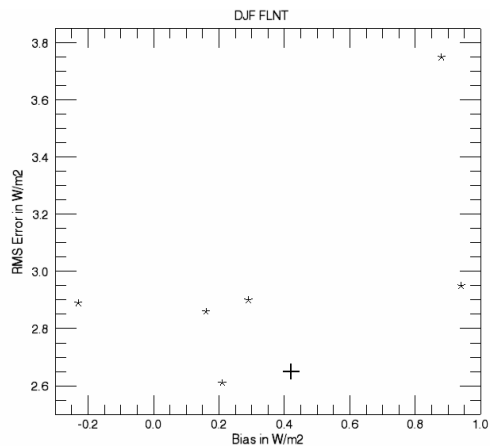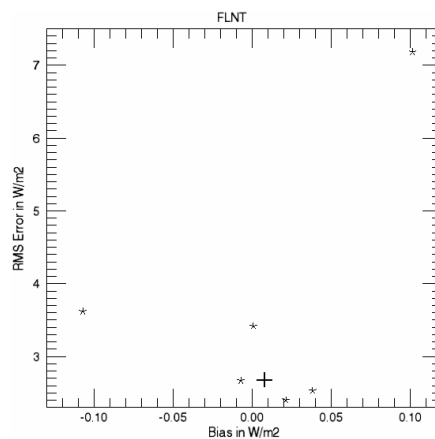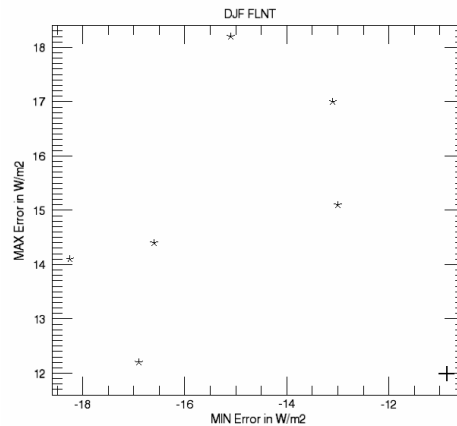

Fig. 11 Ensemble-3 with 6 members. Climate simulations statistics: RMS interpolation error vs. systematic error (Bias) for FLNT (see text). Asterisks – ensemble members, cross – climate NN ensemble results; DJF stands for December-January-February (winter).
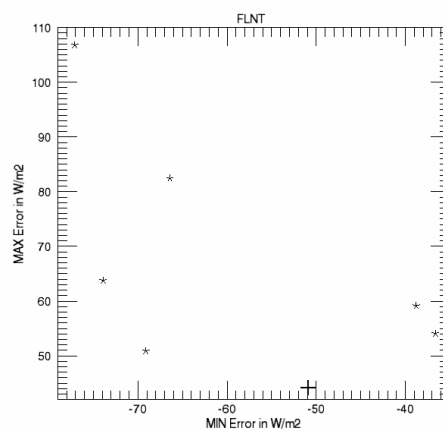

Fig. 14 Ensemble-3 with 6 members. Extreme outliers positions, maximum error vs. minimum error for FLNS (see text) on independent test set. Asterisks – ensemble members, cross – NN ensemble.

ensemble approach, for improving the accuracy of climate simulations, namely for: (a) significantly reducing the systematic and random interpolation error, (b) significantly reducing the magnitudes of the extreme errors or outliers and, (c) in general, significantly reducing the number of large errors. The most important overall result of this study is that the NN emulation ensemble approach provides a positive impact on climate simulation. However, at the next step of this development we have to validate the climate simulation with NN emulations not only against the control simulation with the original LWR, as it is done in this paper, but also against observations.

REFERENCES

[1] K. Funahashi, "On the Approximate Realization of Continuous Mappings by Neural Networks". *Neural Networks*, 2, pp. 183-192, 1989.

[2] V. Krasnopolsky, L. C. Breaker, and W. H. Gemmill, "A neural network as a nonlinear transfer function model for retrieving surface wind speeds from the special sensor microwave imager", *J. Geophys. Res.,* 100, pp. 11033–11045, 1995.

[3] V. M. Krasnopolsky, W. H. Gemmill, and L. C. Breaker, "A multi-parameter empirical ocean algorithm for SSM/I retrievals", *Canadian Journal of Remote Sensing*, 25, pp. 486–503, 1999.

[4] V.M. Krasnopolsky, D.V. Chalikov, and H.L. Tolman, "A neural network technique to improve computational efficiency of numerical oceanic models", *Ocean Modelling,* 4, pp. 363-383, 2002.

[5] V.M. Krasnopolsky, M.S. Fox-Rabinovitz, and D.V. Chalikov, "New Approach to Calculation of Atmospheric Model Physics: Accurate and Fast Neural Network Emulation of Long Wave Radiation in a Climate Model", *Monthly Weather Review*, v. 133, No. 5, pp. 1370-1383, 2005.

[6] V.M. Krasnopolsky and M.S. Fox-Rabinovitz , "Complex Hybrid Models Combining Deterministic and Machine Learning Components for Numerical Climate Modeling and Weather Prediction", *Neural Networks,* 2006, in press

[7] U. Naftaly, N. Intrator, and D. Horn, "Optimal ensemble averaging of neural networks", *Network: Comput. Neural Syst.,* 8, pp. 283-294, 1997

[8] D. Opitz, and R. Maclin, "Popular ensemble methods: an empirical study", *J. Artificial Intelligence Res.,* 11, pp. 169-198, 1999.

[9] S. Hashem, "Optimal linear combination of neural networks", *Neural Networks*, 10, pp. 599-614, 1997.

[10] R. Maclin, and J. Shavlik, "Combining the predictions of multiple classifiers: using competitive learning to initialize neural networks", in *Proceedings of the Eleventh International Conference on Artificial Intelligence*, pp.775-780, Detroit, MI, 1995

[11] W.W. Hsieh, "Nonlinear principle component analysis by neural networks", *Tellus,* 53A, pp. 599-615, 2001

[12] V. M. Krasnopolsky, "Calculation of accurate and stable NN Jacobians using an ensemble approach", *Proceedings of the IJCNN2006, July 16-21, 2006*, Vancouver, BC, Canada, 2006, in press

[13] A. J. C. Sharkey, "On combining artificial neural nets", *Connection Science*, 8, pp. 299-313, 1996.

[14] W.D. Collins, J.K. Hackney, and D.P. Edwards, "A new parameterization for infrared emission and absorption by water vapor in the National Center for Atmospheric Research Community Atmosphere Model", *J. Geophys. Res*., 107 (D22), pp. 1-20, 2002

[15] S.V. Barai and Y. Reich, "Ensemble modeling or selecting the best model: Many could be better than one", *Artificial Intelligence for Engineering Design, Analysis and Manufacturing,* 13, pp. 377-386, 1999